APPENDIX D

STATISTICAL FORMULAE

APPENDIX D. STATISTICAL FORMULAE

D.1. CELL FREQUENCIES

The observed outcome of a case-control study or a cohort study may be depicted in a 2 x 2 table, where a, b, c, and d are cell frequencies.

		ETS Exposed	
		Yes	No
Lung Cancer Present	Yes	a	b
	No	c	d

D.2. CASE-CONTROL STUDIES

The true (but unknown) odds ratio is estimated by the observed odds ratio (OR),

$$OR = ad/bc.$$

A confidence interval on the (true) odds ratio may be calculated from the normal approximation to the distribution of log(OR), the natural logarithm of OR (Woolf, 1955). The variance of log(OR) is estimated by

$$Var(\log(OR)) = 1/a + 1/b + 1/c + 1/d$$

and the standard error by its square root,

$$SE(\log(OR)) = (Var(\log(OR)))^{1/2}.$$

Approximate 90% confidence limits are given by

$$\log(OR) \pm 1.645\ SE(\log(OR)).$$

The value 1.645 is replaced by 1.96 for 95% confidence limits and, in general, by $Z_{\alpha/2}$ for 100(1 - α)% confidence limits. The confidence bounds obtained in this way are sometimes called *logit limits* (Breslow and Day, 1980, p. 134). Significance level (p-value) of a test for effect, i.e., H_o: (true) odds ratio = 1 against the alternative H_a: (true) odds ratio > 1, is the area under the standard normal curve to the right of the value of the *test statistic*, given by log(OR)/SE(log(RR)).

If the (true) odds ratios are assumed to be equal in k studies, then a pooled estimate is calculated from

$$\log(OR(P)) = \Sigma w_i \log(OR)_i / \Sigma w_i$$

where the summations are on i, from 1 to k; OR(P) is the pooled estimate; $\log(OR)_i$ is the logarithm of OR from the i^{th} study; and $w_i = (Var(\log(OR)_i))^{-1}$ is the *weight* of the i^{th} study (Breslow and Day, 1980).

D.3. COHORT STUDIES

The true (but unknown) relative risk is estimated by the observed relative risk (RR),

$$RR = (a/a+c)/(b/b+d).$$

A confidence interval on the (true) relative risk may be calculated from the normal approximation to the distribution of log(RR), using the analogue of Woolf's method referred to above (Katz et al., 1978). The variance of log(RR) is estimated by,

$$Var(\log(RR)) = c/(a^2 + ac) + d/(b^2 + bd)$$

and the standard error by its square root,

$$SE(\log(RR)) = (Var(\log(RR)))^{½}.$$

The remaining calculations follow the description for case-control studies in Section D.2 with "odds ratio" and "OR" replaced by "relative risk" and "RR," respectively. The pooled estimate of relative risk from both case-control and cohort studies is calculated by the same methodology for pooling estimates from case-control studies or from cohort studies separately, i.e., the logarithm of each individual estimate is weighted inversely proportional to its estimated variance (Kleinbaum et al., 1982).